DOCUMENT RESUME

ED 175 918 TM 009 558

AUTHOR Egelston, Richard L.
TITLE Attitude Changes Toward Ten Evaluation Situations As a Result of TAPI Workshops.
INSTITUTION State Univ. of New York, Albany. Educational Research and Service Center.
PUB DATE May 79
NOTE 22p.; For related document, see TM 009 557; Paper presented at the Annual Meeting of the New England Educational Research Organization (Provincetown, Massachusetts, May 1, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Administrator Attitudes; *Adoption (Ideas); *Changing Attitudes; Elementary Secondary Education; Higher Education; Participant Satisfaction; *Supervisory Training; *Teacher Evaluation; *Workshops
IDENTIFIERS *Teacher Appraisal for Improvement

ABSTRACT

Phi Delta Kappa's Teacher Appraisal For Improvement Workshop (TAPI) was examined relative to its effectiveness in changing the attitudes of teachers, administrators, and others toward teacher evaluation. A sample of 315 respondents (mostly administrators), from 13 workshops conducted by two leaders, was given a pre workshop and a post workshop attitude inventory. The inventory consisted of ten situations describing teacher evaluation criteria. The criteria included: unstructured or systematic evaluation by administrators, peers, or students; teacher skill testing; and scores on standardized or objectives-based tests. Respondents were asked to indicate their degree of willingness to implement the situations in their schools on a five-point Likert scale. Repeated measures analysis of variance for each item revealed that the three groups differed significantly on four situations: teachers were less willing to use most evaluations than administrators. While the attitude change for all groups was positive, the magnitude was small. Furthermore, no indication was provided of the long-term effects of TAPI. (A table summarizing the ten approaches to teacher evaluation is appended.) (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Educational Research and Service Center

ED175918

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard L. Igelston

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE UNIVERSITY AT ALBANY
STATE UNIVERSITY OF NEW YORK

Attitude Changes Toward Ten Evaluation Situations As a Result of TAFI Workshops

by

Richard L. Egelston

Presented at the Annual Meeting of

The New England Educational Research Organization

May 1, 1979 at Provincetown, MA

CENTER FOR RESEARCH

STATE UNIVERSITY OF NEW YORK AT ALBANY

ABSTRACT

Phi Delta Kappa's Teacher Appraisal For Improvement Workshop TAFI was examined relative to effectiveness in changing the attitudes of teachers, administrators and others toward teacher evaluation. A sample of 315 respondents from 13 workshops conducted by two leaders was given a pre workshop and a post workshop attitude inventory. The inventory consisted of ten evaluation situations, and respondents were asked to indicate their degree of willingness on a 5 point Likert scale to implement the situations in their schools. Repeated measures ANOVA for each item revealed that the three groups differed significantly on four situations, and that significantly positive attitude changes were affected on seven situations. It was concluded that TAFI was having the desired effect, but that the changes in attitude would probably not result in teacher support for evaluation.

Teacher Appraisal For Improvement (TAFI) is organized as a seven hour workshop designed to make teachers, administrators and others involved with education aware of the variety of ways teaching can be appraised. The emphasis of TAFI is upon the improvement of teaching by the individual teacher. Within each of the eight methodologies discussed the focus of TAFI is upon collecting hard data. Since TAFI is an awareness workshop, little emphasis is placed upon how to create good data collection instruments, although each of the methodologies is presented with at least one example. No attempt is made to describe staff development activities which might be necessary to effect an adoption of any of the methodologies discussed in the workshop. The present study is an attempt to determine if the workshop changes attitudes toward particular evaluation situations.

Background

TAFI was developed by the Center for the Study of Evaluation at UCLA during the middle 1970's. In 1976 an arrangement was made with Phi Delta Kappa to publish the materials, to train leaders for the workshop, and to conduct the workshop. During the past three years 5 leadership training workshops have been held, and about 70 TAFI workshops involving 2,200 participants have been conducted at various locations in the U.S.

Each TAFI workshop is recommended to contain between 30 and 60 participants, although sessions have been as small as 14 and as large as 95 with a single workshop

leader.

A March, 1977 summary of TAFI produced by Phi Delta Kappa provides the following description:

> There are three objectives towards which the workshop is designed: a) to list the major components, advantages, and disadvantages of current methods of teacher appraisal; b) to identify techniques and procedures that can be used to improve these methods; and c) to consider them for application in meeting their own teacher appraisal needs.
>
> Each participant receives a notebook containing information about the various kinds of teacher appraisal systems discussed along with the directions for the various team exercises. Each three member team of participants receives a team notebook containing the team exercises and answers to these exercises. At the conclusion of the workshop the participant's notebook serves as a reference tool on teacher appraisal methods.
>
> Eight techniques and methods that may be useful in the appraisal of teaching are presented. Brief examples of their application are presented and their strengths and weaknesses are outlined. These eight are:
>
> a. Unstructured Observations-(Subjective and idiosyncratic observation of teaching by administrators, peers, or students)
>
> b. Systematic Observations and Ratings-(Use of prespecified procedures for observing teaching and recording descriptions of teaching behaviors: e.g., Flanders Interaction Analysis, teacher skill or characteristics checklists, etc., completed by administrators, peers, or students)
>
> c. Teacher Skill Testing-(The presentation of a teaching-learning situation via a written or verbal description of an audio or video taped presentation and asking teachers to prescribe instructional strategies appropriate to the situation)
>
> d. Omnibus Measures of Student Growth-(The use of standardized tests to determine amount of student growth related to general educational goals and thus to infer quality of teaching)

e. Objectives-based Measures of Student Growth-(Testing students on specific objectives concentrated on during periods ranging from several weeks to a semester)

f. Very Short Measures of Student Growth-(Structuring a teaching-learning situation concentrating on a single objective and observing teaching and learning related to that objective)

g. Teacher Performance Testing Systems-(Structuring a teaching-learning situation concentrating on a single objective and observing teaching and learning related to that objective)

h. Analysis of Typical Student Growth-(The use of regression analysis to identify effective teachers and subsequently the techniques used by teachers whose students perform at or above predicted achievement levels)

In addition, the workshop deals with self vs. other evaluation, process/product evaluation, and various purposes for evaluating teaching.

The workshop contains two phases. In the first phase, the participants are grouped into three-person teams. The teams receive instruction about a particular teacher appraisal method, e.g., ratings. This instruction is provided by means of the participants reading a certain section of their notebooks and/or the workshop leader presenting a brief lecture on the topic and/or a structured discussion of the topic. At the conclusion of the instructional period, the participants complete a team exercise. These exercises provide "hands on experience" in solving realistic problems associated with the exercise. Feedback is then provided as to how the exercise problem should have been solved. There is then a discussion of these answers and the potential applicability and utility of the method for the needs of the participants. This sequence of instruction-exercise-feedback-discussion is repeated for each of the following general kinds of teacher appraisal methods: ratings, teacher skill competency tests, and methods based on student performance. Within each of these general methods, several approaches are considered. For example, within the category of Ratings, the workshop considers ratings by peers, experts,

students, and outsiders in conjunction with various strategies for performing such ratings. These latter strategies range from impressionistic judgements to systematic and qualified data gathering on specific student and teacher actions.

The second phase of the workshop is a simulation of methods considered in the first phase. This involves random assignment of participants to the roles of "teachers,: "students" or "expert observers." Specially prepared packets are distributed to participants to aid them in role playing. A short lesson is actually taught and the results of the various appraisal methods are statistically compared. At the conclusion of the simulation exercise, the participants complete and discuss a questionnaire.

With the increasing emphasis toward accountability in education, many boards of education and district administrators are seeking ways of improving the educational system. Traditionally the manner of effecting change has been through teacher selection and retention rather than through staff development activities. Although there has been a recognition of the need to improve classroom performance by teachers, there was little the district could do to provide assistance in these efforts. TAFI represents one approach which can be used to provide the sensitization within a district to undertake change in a systematic manner. Each of the eight methods is discussed with reference to providing formative evaluation to the teachers who desire appraisal, rather than to provide district officials with summative information for the purposes of assignment, promotion, merit pay or tenure. A careful attempt is made by both the workshop materials and the workshop leaders to avoid endorsing any particular methodology or combination of methodologies. No endorsement of either process evaluation or product evaluation

strategies is made. In fact, one of the initial exercises of the workshop is to provide an opportunity for the participants to indicate a particular orientation toward evaluation, and then to demonstrate the potential conflicts of evaluator and evaluatee who have differing orientations.

Procedures

TAFI workshops have been initiated in two ways; through a request by a school district or another organization to PDK, and by an announcement that one will be held with participants being solicited from a mail campaign. In the former situation the costs of the workshop are generally borne by the requesting organization, and participants contribute very little. In the latter case, the participants contribute the whole cost, although many are subsequently reimbursed most of their expenses.

Thirteen workshops were included for this investigation, of which eight were requested by organizations. Two workshop leaders conducted the set of thirteen with 1 1/2 of the sessions being a team approach. Workshops were held in New York, Vermont, Maine, Connecticut, Ohio, and North Carolina.

Although the workshops were open to anyone who wished to attend, the large majority of participants were school administrators having some role in supervision. A few of the participants were teachers, some were administrative support personnel, some represented higher education, and there were even a few Board of Education members. For the

purpose of this investigation, participants were subdivided into three groups: administrators, teachers, and others. A person was classified as an administrator if the role was in a direct line of supervision. These roles included department chairpersons (who may also be teaching), principals, assistant superintendents or directors of instruction, and chief school administrators. No distinction was made between public and private schools or any designation between K and 12. No verification was attempted for determining if a person had a supervisory role, and some misclassifications may have occurred.

A participant was considered as a teacher if their part-time or full-time role was in teaching, and they did not supervise other teachers. Again this category contained public and private school teachers between K and 12. The remaining category, Others, included administrators, faculty, and graduate students in higher education, lay people, and non-supervisory administrators within central offices.

All participants were encouraged to remain for the full workshop, and to participate in all of the workshop activities. However, some people arrived late at each workshop and consequently were not able to complete the pre-workshop inventory. Others in each session left early and failed to complete the post-workshop inventory. Those participants who completed each inventory were asked to indicate their job title, and to include the last four digits of their home telephone number, but not to provide any other

personal identifying information. Those participants whose inventories could be paired provided the groups for this investigation. These groups totaled 315 out of the larger set of 573 participants who completed at least one inventory. Within the study there were 241 administrators, 17 teachers, and 57 others. Unusable responses slightly reduced these group sizes for some analyses.

The attitude inventory contained ten short descriptions of evaluation situations which could be employed in a building or a district for the purposes of providing information on teacher appraisal (see Table 1). Directions to the respondent (see Table 2) indicated that each situation was to be considered on its own merit, and that the respondent should indicate a degree of willingness to implement it within the school building or district. As can be seen from Table 2, each strategy was rated on a five position Likert scale with the category of very unwilling coded 1 to the category of very willing coded 5. The scale was considered an equal appearing interval scale (Tuckman, 1972; p. 157) which led to the adoption of parametric analyses. No attempt was made to analyze the selection of combined strategies.

Insert Tables 1 and 2 about here

Analyses of the data were carried out with repeated measures analyses of variance for each evaluation situation. The within subjects factor was the pre and post workshop attitude response, and the between groups factor was the classification of job role. Only one of the ten situations (#3) had both responses from all 315 respondents, however the smallest number of paired responses was 310 (situation number 8). All analyses were carried out with the aid of the BMPD2V program on the SUNY at Albany Computer System.

Results

Table 3 reports the means, standard deviations and approximate cell sizes for each of the ten evaluation situations relative to both the pre workshop and the post workshop attitudes. For example, the mean attitude of administrators on situation B was 3.74 on the initial measurement and 3.90 on the final measurement indicating that the workshop had a small influence on using systematic ratings with known standards of performance. Teachers also increased their willingness to use this situation, and had a slightly larger increase than administrators. All three groups indicated a greater degree of willingness to use situation B as a result of the workshop.

Insert Table 3 about here

Evaluation situations A and C were the only ones to show an overall decrease in the willingness of participants to implement them. Situation A dropped .04 with most of

the change being expressed by the Others group. Situation C had a smaller decrease, although the decrease was consistent across all three groups.

Situation B produced the highest ratings of all ten situations, and could be interpreted as the only one in which the willingness of the participants to use it was substantially above the neutral position. However, with the exception of situations A and C, all three groups were more willing to use the situations as a result of the workshop. This finding does not mean that the groups actually were willing to use them (with the possible exception of situation B) since most of the overall means were at or below the neutral position. Either a single evaluation methodology is unacceptable, or there is a basic resistance to publically admitting improvement is needed.

Teachers were less willing to use most evaluation methods than administrators. Perhaps this result could be attributed to a feeling of being victimized, or threatened by evaluation. In those situations where the teachers were more willing (situations A and E), the difference in means was .06 or less.

Due to space limitations only the ANOVA Table for situation J is presented in full form. The main effects of groups and pre-post are both significantly different from zero at the .05 level (see Table 4). Reference to the means presented in Table 3 indicates that all three groups gained in willingness to use the situation as a result of

the workshop. The interaction of group and time of measurement was not significant.

Insert Table 4 about here

Table 5 summarizes the results of all ten ANOVA's by reporting the values for F with an indication of their significance. The last four situations(G,H,I,J) were the only ones to produce a significant difference among groups. With two exceptions, teachers were the least willing to utilize the ten situations. Administrators and Others were split on being the intermediate and the most willing groups. This pattern of differences between groups indicates that they should not be pooled for analysis.

Insert Table 5 about here

Table 5 also indicates that seven of the ten situations produced a significant attitude change as a result of the workshop. In each of these situations the change was toward a greater willingness to use the situation (see Table 3). The sizes of the gains were small, ranging from .09 for situation B to .23 for situation J. Situation D produced a gain of .05, but situation A had a decline of .02 and situation C a decline of .01.

Discussion

The consistant gains across situations indicates that the TAFI workshop is capable of producing significant positive changes in attitudes toward the appraisal of teaching. Whether the increased acceptance is due to a

new familiarity of techniques is uncertain, but remains one possible explanation of the findings. Doubtless, some of the situations described in the attitude instrument were unfamiliar on the pretest. Several others may have been given little consideration up to that point, and a hasty judgement produced. If this conjecture were true, it might indicate an initially cautious attitude on the part of participants, with a subsequent liberalizing of attitude when they understand the techniques could be used in a fair manner. Since a great deal of discussion within TAFI concerns the understanding between the evaluator and evaluatee about what variables are going to be assessed, and what criteria are going to be employed following the assessment, some distrust of the methodology may have been eliminated.

The less than enthusiastic support of teachers and administrators toward any of the evaluation situations raises questions about the degree of change in behavior brought about by the awarness of previously unfamiliar strategies. While the change in attitudes were in the proper direction, the magnitude was small. Furthermore, no indication of whether the attitudinal change was long lasting was attempted.

For those participants (particularly teachers) who do not desire to be appraised, TAFI provides a great deal of ammunition for resisting evaluation. For those who have the ego-strength to undergo appraisal, the workshop provides new-frameworks for collecting data as well as

interpreting its meaning. Those who are committed to the improvement of instruction will have an increased set of tools for evaluation, but will probably feel discomforted by the low level of acceptance of the teacher groups.

Although the TAFI materials are highly structured, and the leader is requested to read much of the material aloud during the sessions, the developers have produced a series of marginal notes and important points to cover at each stage of the workshop. The leader who appears to follow the text less and can direct discussion without obvious breaks in the flow will experience more enthusiasm on the part of the participants. One interesting research question to be raised lies in the effect of the workshop leader. The standardized format attempts to reduce the variation among TAFI leaders, but the standardization may have a negative effect upon the participants. Perhaps the leadership training should incorporate some practice in delivery of TAFI. This TAFI leader has undertaken a series of changes in workshop organization, enough so that each session is somewhat different from all of the others. Although the changes appear to have an influence on dropout rate, the influence on attitudes toward the evaluation situations is unknown.

Conclusion

TAFI has the capability of producing a change in attitudes toward the adoption of a variety of evaluation strategies. These changes are generally in a positive direction, and tend to be relatively uniform across groups.

However, the post-workshop attitudes are still neutral at best. Therefore, more work is needed on the evaluation process. Perhaps this could take the form of an extended workshop over two or three days in which it would be possible to jointly develop instruments which would more appropriately measure the variables the individual teacher and administrator believes is important. Perhaps another avenue would be follow-up with staff development activities. TAFI can get some participants all fired up, but cannot tell them how to follow through.

References

Fink, A., Klein, S. & Burry, J. Teacher Appraisal for Improvement. Bloomington, IN: Phi Delta Kappa, 1976. (Leaders Manual, Participants Notebook, Team Materials).

Phi Delta Kappa. The teacher appraisal for improvement (TAFI) workshop. Bloomington, IN: Author, 1977.

Tuckman, B., Conducting Educational Research. New York Harcourt, Brace, Jovanovich, 1972.

Table 1

PRACTICAL USES OF TEACHER APPRAISAL APPROACHES

A. The principal and/or other supervisory personnel write a report of the teacher's effectiveness. This report is based on personal acquaintance with the teacher, comments by students, parents, and colleagues, and informal classroom observation.

B. The principal and/or other supervisory personnel systematically observe and rate the teacher using a rating form that lists important characteristics of good teaching. The teacher is given a score showing the extent to which he/she exhibits these characteristics in terms of a predetermined school or district standard of effectiveness.

C. The teacher's peers systematically observe and rate him/her on the extent to which he/she exhibits important characteristics of good teaching. The teacher must meet a predetermined school or district standard of effectiveness.

D. The teacher's students use a rating form to judge the extent to which the teacher exhibits important characteristics of good teaching. The teacher must meet a predetermined school or district standard of effectiveness.

E. The teacher is given written descriptions and/or shown films of typical classroom problems. The teacher's effectiveness is judged on the basis of the quality of the answers.

F. Periodically, the teacher is provided with an instructional objective, a sample test item measuring that objective, and information about the content it covers. A small group of students is assigned to that teacher randomly (to balance abilities), and is instructed by the teacher on the objective for one to ten lessons. After instruction, the students are tested on the objective. Teacher effectiveness is determined on the basis of how well the students achieved the objective.

G. The teacher instructs his/her students on relevant instructional objectives for a period of time ranging from several weeks to one semester. At the end of each instructional period, the students are tested on their attainment of those objectives. The level of student achievement provides the basis for appraising teacher effectiveness.

H. Nationally or locally normed standardized tests are administered to students at the end of each school year. Teacher effectiveness is appraised on the basis of the extent to which student performance conforms to the test's established norms.

I. Nationally or locally normed standardized tests are administered to students to determine how much they increase their learning over time. The amount of desired increase is determined in advance by relevant school personnel, including teachers. Teacher effectiveness is appraised on the basis of the extent to which students achieve the desired increase.

J. The students in each grade or subject matter area are tested at the beginning and end of each school semester or year. Student scores are statistically analyzed to determine which class grows more or less than classes of comparable ability. Teacher appraisal is based on the proportion of "gainers" to "losers."

Table 2

Rating Sheet for Ten Evaluation Situations

Job Title: _____ Tel # _____

Alternate Approaches to Teacher Appraisal: Questionnaire

Ten major approaches to appraising teacher effectiveness are described on page R-1 of your participant book. Please evaluate each approach according to the extent to which you would be willing to implement it in your school or school district. Evaluate each approach on its own merits. Do not assume that a given approach would be used with some other approach.

The following scale has been provided to assist you in making your evaluations:

1 = very unwilling
2 = not willing
3 = neutral
4 = willing
5 = very willing

Next to the letter of each approach, circle the number of the choice that comes closest to describing the extent to which you would or would not be willing to use this one approach in your school. Remember, the higher the scale value chosen the more you would be willing to use the approach.

Approach	Rating	Approach	Rating
A	1 2 3 4 5	F	1 2 3 4 5
B	1 2 3 4 5	G	1 2 3 4 5
C	1 2 3 4 5	H	1 2 3 4 5
D	1 2 3 4 5	I	1 2 3 4 5
E	1 2 3 4 5	J	1 2 3 4 5

After you have finished rating each approach on its own merits, select the method of Teacher Appraisal that you would be most willing to implement in your school or school district. The method you select may involve one of the approaches listed above, or any practical combination of these approaches. Indicate your selection by circling the appropriate letter(s).

A B C D E F G H I J

Table 3 Group Means and Standard Deviations for Ten Evaluation Situations

Situation	Statistic	Pre Workshop			Post Workshop		
		Admin.	Tchrs.	Others	Admin.	Tchrs.	Others
1-10	Approx n	237-241	16-17	56-57	237-241	16-17	56-57
A	mean	3.14	3.18	3.10	3.13	3.18	2.86
	st. dev.	1.24	1.19	1.22	1.22	1.13	1.27
B	mean	3.74	3.19	3.58	3.90	3.44	3.77
	st. dev.	1.07	1.28	1.07	.92	.96	.87
C	mean	2.77	2.82	2.95	2.75	2.65	2.93
	st. dev.	1.12	1.38	1.17	1.04	1.32	1.07
D	mean	2.42	2.35	2.60	2.49	2.47	2.86
	st. dev.	1.08	1.00	1.07	1.09	1.07	1.08
E	mean	2.04	2.06	1.91	2.38	2.44	2.30
	st. dev.	1.02	1.00	.96	.99	1.09	.91
F	mean	2.55	2.12	2.56	2.91	2.82	3.00
	st. dev.	1.11	1.05	1.05	1.04	1.01	.94
G	mean	2.97	2.35	3.05	3.29	2.71	3.52
	st. dev.	1.07	1.06	1.41	.92	.85	.87
H	mean	2.29	1.65	2.21	2.63	1.94	2.80
	st. dev.	1.02	.86	1.40	1.05	1.03	1.13
I	mean	2.86	2.00	2.77	3.05	2.47	3.18
	st. dev.	1.11	1.06	1.41	.99	1.37	1.01
J	mean	2.61	1.65	2.45	3.02	2.00	3.14
	st. dev.	1.15	1.17	1.41	1.09	1.22	1.15

Table 4 ANOVA Table for Situation J

Source	SS	df	MS	F	p
Group (A)	31.44	2	15.72	8.59	$< .05$
S/A	565.52	309	1.83		
Pre-Post (B)	13.17	1	13.17	15.61	$< .05$
AB	1.96	2	.98	1.16	
BxS/A	260.77	309	.84	ns	
Total	1366.86	623			

Table 5 Summary of ANOVA Tables for Three Groups Responding With Attitudes Toward Ten Evaluation Situations at the Beginning and End of TAFI Workshops

Situation	F Values		
	Between Groups	Pre-Post	Interaction
1	.82	.28	.33
2	2.92	4.42*	.10
3	.83	.49	.18
4	2.10	2.18	.74
5	.39	12.20*	.06
6	.90	17.45*	.60
7	5.15*	12.23*	.32
8	4.61*	13.17*	1.04
9	5.01*	10.26*	1.01
10	8.59*	15.61*	1.16

* $p < .05$